VI International Conference on Computational Methods for Coupled Problems in Science and Engineering
COUPLED PROBLEMS 2015
B. Schrefler, E. Oñate and M. Papadrakakis(Eds)

ON THE TERM AND CONCEPTS OF NUMERICAL MODEL VALIDATION IN GEOSCIENTIFIC APPLICATIONS

Uwe-Jens Görke*, Thomas Nagel* and Olaf Kolditz*,†

* Department of Environmental Informatics
Helmholtz Centre for Environmental Research - UFZ
Permoserstr. 15, D-04318 Leipzig, Germany
e-mail: uwe-jens.goerke@ufz.de - Web page: http://www.ufz.de

† University of Technology Dresden, Germany

Key words: Coupled Problems, Model Development, Numerical Methods, Model Confirmation, Inverse Problems

Abstract. Modeling and numerical simulation of the coupled physical and chemical processes observed in the subsurface are the only options for long-term analyses of complex geological systems. This contribution discusses some more general aspects of the (dynamic) process modeling for geoscientific applications including reflections about the slightly different understanding of the terms *model* and *model validation* in different scientific communities, and about the term and methods of *model calibration* in the geoscientifc context. Starting from the analysis of observations of a certain part of the perceived reality, the process of model development comprises the establishment of the physical model characterizing relevant processes in a problem-oriented manner, and subsequently the mathematical and numerical models. Considering the steps of idealization and approximation in the course of model development, Oreskes et al. [1] state that process and numerical models can neither be verified nor validated in general. Rather the adequacy of models with specific assumptions and parameterizations made during model set-up can be confirmed. If the adequacy of process models with observations can be confirmed using lab as well as field tests and process monitoring, the adequacy of numerical models can be confirmed using numerical benchmarking and code comparison. Model parameters are intrinsic elements of process and numerical models, in particular constitutive parameters. As they are often not directly measurable, they have to be established by solving inverse problems based on an optimal numerical adaptation of observation results. In addition, numerical uncertainty analyses should be an obligatory part of numerical studies for critical real world applications.

1 INTRODUCTION

Modeling and numerical simulation of the coupled physical and chemical processes observed in the subsurface are the only options for long-term analyses of complex geological systems. The numerical simulation of geotechnological processes (*dynamic modeling*) requires the existence of models that describe the considered problems as adequately as possible. Within this context, it has to be considered that models represent approximations and idealizations of the considered part of the perceived reality, and will map the real world by approximation. During the process of model development, certain relevant aspects will be covered sufficiently accurate, whereas other details, which can be considered as irrelevant for the specific problem definition, can be neglected. In this approximate sense, models are qualified to enable predictions of the behavior of real-world processes, which are not (yet) open for efficient empirical measuring procedures due to exceptional long time scales to be considered (e.g., within the context of geological waste deposition) or due to the general difficult access to local *in situ* measurements in the subsurface. However, absolute exact predictions can never be expected based on process modeling. The model quality, i.e., the degree of conformance of models with the part of the perceived reality they describe, and thus the reliability of model predictions, depends on many factors. Within this context, the quality as well as the spatial and time density of measured data characterizing the observation area plays a crucial role. These data are necessary for parameterization (i.e., calibration; parameter identification; inverse modeling) and confirmation of the developed models.

2 MODELS WITHIN A GEOSCIENTIFIC CONTEXT – AN ATTEMPT OF DISAMBIGUATION

Discussing about *models* in a geoscientifc context, different scientific communities (e.g., geologists; geophysicists; experts in continuum mechanics; engineers) sometimes use this term based on slightly different associations. Thus, in the following we attempt to provide clear definitions of the term *model*:

- **Spatial and/or structural models** usually will be summarized using the term *(static) reservoir model*, and comprise all empirically ascertainable information regarding geometry, geological structure, basic material characteristics (e.g., porosity, intrinsic permeability, thermal and mechanical parameters), and reservoir conditions (e.g., pressure, temperature) of the considered area.

- **Process models** represent *mathematical models* based on physically founded assumptions about coupled physical, geochemical, geoelectrical, and (micro-)biological processes observed in real-world applications. Basically, process models include balance laws and constitutive relations. More detailed reflections regarding process modeling will be discussed in Sec. 3 of this contribution.

- **Numerical models:** The field equations based on balance laws including relevant constitutive relations can be solved analytically just in exceptional cases. Corresponding specific problems frequently will be defined in terms of benchmarks for model and software confirmation. Usually, the solution of field equations describing a real-world problem requires the use of numerical methods. For that, the local formulation of balance laws in terms of differential equations has to be transformed into a global integral formulation. In general, the resulting system of equations is a nonlinear one and contains time derivatives of the primary variables to be calculated. As numerical methods do not provide spatially and temporally continuous solutions, but rather solutions in discrete points of the given solution space, the system of field equations including constitutive relations has to be discretized in space and time within the context of incremental-iterative approximation procedures. In addition, nonlinear systems have to be linearized. The resulting so-called *initial-boundary value problem* can be solved numerically in discrete locations at discrete points of time. As spatial discretization approaches, the methods of finite differences (FDM), finite volumes (FVM), and finite elements (FEM) are widely excepted means of choice. Within this context, the term *numerical model* characterizes either
  - numerical methods and algorithms necessary for computer-based simulations of the process model, or
  - the entire data set necessary for the numerical simulation of a specific problem (parameterized and spatially discretized structural model; boundary and initial conditions; parameters controlling the simulation procedure).

3 PROCESS MODELING

The process modeling of physical, geochemical, and (micro-)biological processes in the subsurface is based on mathematical theories enabling the description of spatially and temporally evolving processes of the perceived reality in terms of differential equations or systems of differential equations. Developing the relevant mathematical apparatus, we make use of physically, chemically, and biologically founded assumptions.

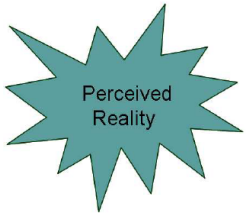

Figure 1: Origin of the development of process models

Thus, based on reasonable abstractions, the process model to be defined can be formulated to be sufficiently complex in order to cover all relevant aspects of the specific

real-world application but at the same time to be sufficiently simple in order to enable its numerical solution. Observations of the interesting, problem-specific part of the perceived reality serve as the nucleus of the development of process models (cf. Fig. 1). A first step in this procedure is the establishment of physical models adequately mapping the perceived reality based on decisions to neglect those processes that are not relevant for the considered application (cf. Fig. 2). This kind of idealization is problem-dependent, and thus, different problems related to the same observation area may result in different physical models (e.g., if a shallow subsurface area is designated for groundwater remediation activities or for use as geothermal reservoir).

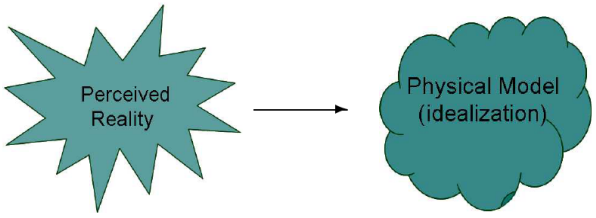

Figure 2: Generation of a physical model

The second step in the definition of process models is the development of a mathematical framework describing the physical model in an approximate manner, and consisting of balance laws and constitutive relations in terms of algebraic, differential or integral equations (cf. Fig. 3).

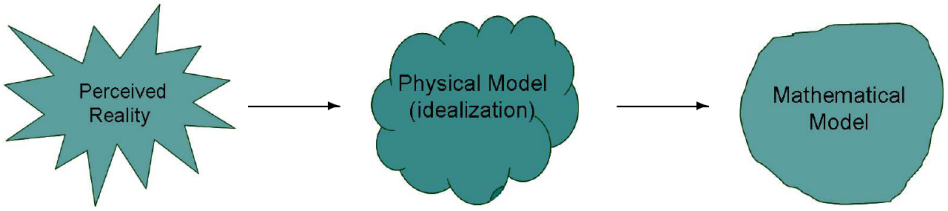

Figure 3: Generation of a mathematical model

As a matter of principle, processes of the nature can be characterized using balance laws related to fundamental state variables. Those are primarily the balances of mass, linear momentum, moment of momentum, and energy. Balance laws are independent from the real materials under consideration, and can be formulated locally in terms of problem-oriented (systems of) differential equations based on convenient assumptions. Within this context, they are valid for each location of the observation area, and are usually called *field equations*.

Observations of the nature show that certain physical variables may evolve in different manner dependent on the specific materials involved in the considered application (e.g., various pictures of deformation of a solid body of the same geometry and subjected to the same external loading, but consisting of different materials). This causes the necessity of the definition of additional equations, the so-called *constitutive relations* characterizing material-dependent dependencies between state variables. Equations of state for fluids, stress-strain relations for solid bodies, and kinetic (reaction) models for chemical reactions are typical examples for constitutive relations. For mathematical reasons, the definition of constitutive relations is necessary in order to define a closed system of equations including balance laws and constitutive relations. The pure system of field equations resulting exclusively from the balance laws contains more unknowns as equations, and can thus not be solved in a unique manner.

4 MODEL CONFIRMATION

The procedure of the development of process models illustrated in Sec. 3 shows that models are always idealizations and approximations of a considered part of the perceived reality reflecting specific assumptions and parameterizations made for the model set-up. In order to assure the usability of a process model and the corresponding numerical model, it has to be shown that it adequately reflects real-world processes. In the literature, this process is called *model validation* or even *model verification* – but are these terms justified in their absolute, literal meaning? According to Oreskes et al. [1]

- **Model verification** "... means an assertion or establishment of truth." and "To say that a model is verified is to say that its truth has been demonstrated, which implies its reliability as a basis for decision-making."
- **Model validation** "... does not necessarily denote an establishment of truth... Rather, it denotes the establishment of legitimacy..." and "... a model that does not contain known or detectable flaws and is internally consistent can be said to be valid."

Considering the steps of idealization and approximation in the course of model development, and the open character of models, Oreskes et al. [1] state that process and numerical models can neither be verified (establishment of the truth of the model) because

- "... models... are never closed systems" (but truth can be demonstrated only for closed systems),
- models "... require input parameters that are incompletely known",
- data sets for the parameterization (i.e., calibration) of process models are always incomplete (e.g., due to the availability of only a few locally determined parameters from core tests considering a model of a whole geological reservoir), and

- often used phenomenological constitutive theories are characterized by loss of information on real physical scale (e.g., use of macroscale models not being based on real microscale behavior),

nor validated (establishing the legitimacy of a model) in general because

- model results depend on assumptions required for model development and on parameterization, and
- models valid for one mapping of the reality may not be valid for another one (e.g., material behavior dependent on temperature).

If the terms model verification and model validation are unfavorable in their literal meaning, how else the usability of process models can be assured? For this purpose, Oreskes et al. [1] introduce the term *model calibration* stating: "...science requires that empirical observations be framed as deductive consequences of a general theory... If these observations can be shown to be true, the theory... is *confirmed*...", and "...confirming observations do not demonstrate the veracity of a model..., they only support its probability." Thus, rather the adequacy of models with specific assumptions and parameterizations made during model set-up can be confirmed, not their (general) truth. Within this context, model parameterization is performed using well-established approaches of model calibration. If the adequacy of process models with observations can be confirmed using lab as well as field tests and process monitoring, the adequacy of numerical models can be confirmed using numerical benchmarking (e.g., providing analytical solutions) and code comparison for more complex systems (see Class et al. [2], Kolditz et al. [3], and others).

Figure 4: Direct modeling: simulation of real-world problems using parameterized process and numerical models

5 MODEL CALIBRATION AS AN OPTIMIZATION PROBLEM

Model parameters, in particular constitutive parameters, are constituent, intrinsic elements of process and numerical models. Parameterized models are used for simulations in order to analyze and/or predict the evolution of various processes in real-world applications, thus, solving a *direct problem* (cf. Fig. 4). Within this context, model parameters

can be considered as *reason* for the specific evolution of relevant primary variables (e.g., fluid pressure, deformation of solids, temperature(s), concentration(s)).

As model parameters are often not directly measurable, their identification is based on the analysis of their *effects* onto measurable field variables. This process of model calibration (i.e., model parameterization) requires the solution of an *inverse problem* based on an optimal numerical adaptation of observation results (cf. Fig. 5)

Due to the lack of information about the perceived reality (i.e., observation results are available only for discrete points of the field problem to be solved for real-world applications), the process of identification of model parameters is an ill-posed problem being source of the non-uniqueness of the solution, and potentially suffering from instabilities of the solution.

Figure 5: Inverse modeling: optimal numerical adaptation of observations from lab and field experiments, and from field exploration

As the operator mapping model parameters to the considered primary field variables is usually of implicit, non-linear character with unknown mathematical structure, its explicit closed-form inversion approves to be impossible in this case. Consequently, the ill-posed inverse problem of model calibration in general results in the solution of a non-linear optimization problem: The parameters have to be estimated in such a way that an appropriately defined objective function approaches its minimum (for an overview see Bruhns and Andig [4], and others).

5.1 Objective Function

A model function will be defined, which characterizes an arbitrary field variable y depending on a vector of variables $\mathbf{x}$ as well as on a set (vector) of model parameters $\mathbf{c}$

$$y \;=\; y(\mathbf{x}, \mathbf{c})\,. \tag{1}$$

In order to analyze a deformation problem, the corresponding model function can be, for instance, constituted by the displacement field depending on stresses, internal variables and material parameters.

The calibration of constitutive models is aimed at the determination of parameters realizing a sufficiently accurate approximation of measured discrete data $\hat{y}_i$ representing defined conditions for variables $\mathbf{x}_i$. The corresponding parameter set, which is in this

narrower sense an optimal one, and thus the best approximation of measured data, can be considered as determined if a least squares norm approaches its minimum

$$\frac{1}{2}\sum_{i=1}^{n}\left[\hat{y}_i - y(\mathbf{x}_i,\mathbf{c})\right]^2 \quad \longrightarrow \quad \min\,. \tag{2}$$

Based on the definition of a vector of residuals $\mathbf{r}$ between measured and calculated values

$$\mathbf{r}(\mathbf{c}) = \left\{r_i(\mathbf{c})\right\} \qquad \text{mit} \qquad r_i(\mathbf{c}) = \hat{y}_i - y(\mathbf{x}_i,\mathbf{c}) \tag{3}$$

the objective function Φ can be formulated as the following least squares norm:

$$\Phi(\mathbf{c}) = \frac{1}{2}\,\mathbf{r}^{\mathrm{T}}(\mathbf{c})\,\mathbf{r}(\mathbf{c}) = \frac{1}{2}\sum_{i=1}^{n} r_i^2(\mathbf{c}) \quad \longrightarrow \quad \min\,. \tag{4}$$

The necessary optimality criterion

$$\nabla\Phi(\mathbf{c}^*) = \mathbf{0} \tag{5}$$

results for models representing linear functions of the parameters $\mathbf{c}$ in a system of linear algebraic equations with the coordinates of the optimal parameter vector $\mathbf{c}^*$ as primary variables. This system is known as *normal equation*

$$\widetilde{\mathbf{J}}^{\mathrm{T}}\,\widetilde{\mathbf{J}}\,\mathbf{c}^* = \widetilde{\mathbf{J}}^{\mathrm{T}}\mathbf{r}\,. \tag{6}$$

Here, $\widetilde{\mathbf{J}}$ represents the Jacobian matrix comprising the first derivatives of the residuals with respect to the material parameters.

5.2 Optimization Procedure

If the identification of model parameters is based on the analysis of homogeneous fields of primary variables, only the initial value problem has to be solved to obtain the value of the objective function to be minimized (e.g., using results from uni- or triaxial compression tests). Apart from deterministic (gradient-based) optimization methods in this case special applications of stochastic and/or evolutional (gradientless) approaches are reported in the literature even for quite complicated material models (see e.g. Harth et al. [5]).

Analyzing experiments on samples with an inhomogeneous distribution of primary variables the amount of information for a reliable identification process is distinctly higher. In this case, the objective function has to be calculated solving a complex initial-boundary value problem, for what usually FEM approaches are used. Thus, due to the time-consuming solution of the direct problem stochastic methods are not suitable for the analysis of inhomogeneous physical, geochemical, and/or (micro-)biological states but deterministic approaches are well-tried for this case in engineering sciences. Regarding

their mathematical basics we exemplarily refer to the monographs of Beck and Arnold [6], Dennis and Schnabel [7], Nocedal and Wright [8] or Rao [9]. A crucial step to be performed using deterministic optimization approaches is the so-called *sensitivity analysis*, i.e., the calculation of the Jacobian matrix of first derivatives of the objective function (cf. Eq. 6). Common methods to perform the sensitivity analysis are analytical calculations if exclusively explicit functional dependencies occur, pure numerical approaches using finite differences, or the direct differentiation (semi-analytical methods).

Considering the modeling of geoscientific real-world problems, the so-called *history matching* is well-established for the parameterization of complete static site models. Within the context, model parameters will be adapted performing simulations of the considered problem with known process parameters (e.g., measured borehole pressures and/or temperatures).

6 CONCLUSIONS

The above illustrated procedure of the development of process and numerical models makes clear that at each stage of model definition certain approximations of the real world are implemented. Thus, modeling and numerical simulation can provide a preview of trends of the behavior of the studied systems in accordance to the defined assumptions and preconditions, but no closed, accurate, deterministic predictions can be expected (modeling is not reality). Consequently, formulations in regulatory documents establishing rules for the implementation of (geo-)technologies have to reflect this character that is inherent to modeling: it is not reasonable to claim exact long-term guarantees, but it is rather useful to claim a statistically founded estimation of trends. Nevertheless, it is out of question that modeling and numerical simulation is the only option for long-term analyses of complex systems, and for providing prognoses in the range of reliable probabilities, if empirical data are difficult to access in lab and/or field experiments. Climate simulations, weather forecast, and geoscientific applications as considered here are exemplary for that. Within this context, it is of vital importance to know what can be expected from modeling, how results of numerical simulations have to be interpreted, and to keep limitations in mind.

The following aspects are of great importance with respect to reliable process simulations:

- Definition of the processes and subprocesses that are relevant for the specific problem to be solved in order to establish a system of field equations, which is complex enough to cover the relevant system behavior but simple enough to ensure an efficient and stable solution process.
- Formulation of problem-dependent specific expressions of the balance laws and the constitutive relations according to the latest state of research.
- Model confirmation using standardized benchmarks and code comparison.

- Extensive provision of data from lab and field tests for an improved process understanding and for the model calibration.
- If measured data are not available for all local points of the spatially discretized observation area (this is the usual case for geoscientific real-world applications), a stochastically based parameter space has to be determined based on a statistically sufficient number of different numerical realizations.
- An uncertainty analysis is advised based on this parameter space, and results of the process simulation will be provided with certain likelihood.

These aspects have to be reconsidered, if the process and numerical models undergo changes in the observation area (e.g., enlargement or reduction in one of the geometric dimensions), and/or in the initial, boundary and/or process conditions, or if the specific problem modifies (e.g., interest in hydro-mechanical coupled processes instead of pure hydrological processes). The motivation behind is that each modeling process includes the definition of problem-oriented input data, and thus, provides specific, problem-oriented results.

As modeling cannot provide closed, deterministic predictions, but rather a preview of trends regarding how the studied systems may behave under the defined assumptions and conditions, the solution of inverse problems in geoscientific applications deserves more attention. Within this context, an extensive provision of lab and field test data for an enhanced process understanding combined with a more reliable and documented model calibration are required to improve the adequacy of static and dynamic models. In addition, numerical uncertainty analyses should be an obligatory part of numerical studies for critical real world applications.

REFERENCES

[1] Oreskes, N., Shrader-Frechette, K. and Belitz, K. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science* (1994) **263**:641–646.

[2] Class, H., Ebigbo, A., Helmig, R., Dahle, H.K., Nordbotten, J.M., Celia, M.A. et al. A benchmark study on problems related to CO_2 storage in geologic formations. *Comput. Geosci.* (2009) **13**:409–434.

[3] Kolditz, O., Görke, U.-J., Shao, H. and Wang, W. (eds.). *Thermo-hydro-mechanical-chemical processes in porous media: benchmarks and examples*. Lecture Notes in Computational Science and Engineering 86, Springer (2012).

[4] Bruhns, O. and Anding, D.K. On the simultaneous estimation of model parameters used in constitutive laws for inelastic material behaviour. *Int. J. Plast.* (1999) **15**:1311–1340.

[5] Harth, T., Schwan, S., Lehn, J. and Kollmann F.G. Identification of material parameters for inelastic constitutive models: statistical analysis and design of experiments. *Int. J. Plast.* (2004) **20**:1403–1440.

[6] Beck, J.V. and Arnold, K.J. *Parameter estimation in Engineering and Science*. Wiley, New York (1977).

[7] Dennis, J.E. and Schnabel, R.B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall Inc., Englewood Cliffs, New Jersey (1983).

[8] Nocedal, J. and Wright, S.J. *Numerical Optimization*. Springer (1999).

[9] Rao, S.S. *Engineering Optimization: Theory and Practice*. 3rd ed., Wiley, New York (1996).